(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number
WO 01/89221 A1

(51) International Patent Classification[7]: H04N 7/18

(21) International Application Number: PCT/US01/16289

(22) International Filing Date: 18 May 2001 (18.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/205,942 18 May 2000 (18.05.2000) US
60/254,453 7 December 2000 (07.12.2000) US

(71) Applicant: **iMOVE INC.** [US/US]; 33 NW First Ave., Suite 1, Portland, OR 97035 (US).

(72) Inventors: **WHITE, Patrick**; 3045 SW 187th Ave, Aloha, OR 97006 (US). **HUNT, Brian**; 2833 NE Everett St., Portland, OR 97232 (US). **RIPLEY, David, G.**; 1436 NW Benfield Drive, Portland, OR 97229 (US).

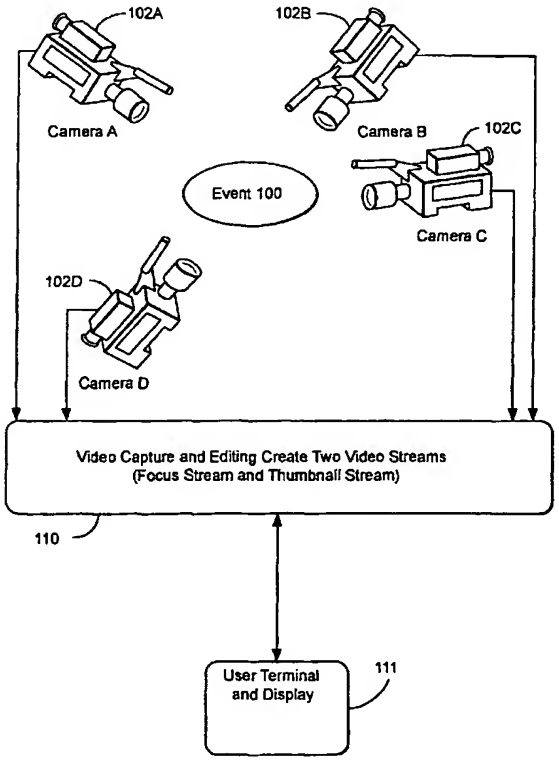
(74) Agent: **GALBI, Elmer**; 13314 Vermeer Drive, Lake Oswego, OR 97035 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,

[Continued on next page]

(54) Title: MULTIPLE CAMERA VIDEO SYSTEM WHICH DISPLAYS SELECTED IMAGES



(57) Abstract: Multiple streams of data are streamed to a user's terminal (111) with images from different cameras (102A, 102B, 102C, 102D). Low resolution thumbnail images tell the user what image streams are available. A focus stream provides high resolution images from a selected camera. A user can switch the focus stream to another stream by clicking on the associated thumbnail (110). The users can also be provided with a thumbnail of panoramic image. Other data streams sent to the user can contain (a) audio data, (b) interactivity markup data which describes regions of the image which provide interactivity opportunities such as hotspots, (c) presentation markup data which defines how data is presented on the user's screen, (d) telemetry data stream which can be used for various statistical data. One data stream contains a low quality base image for each data stream which can be enhanced to form a high resolution focus image.

WO 01/89221 A1

IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Multiple Camera Video System Which Displays Selected Images

Field of the Invention:

The present invention relates to transmitting video information and more particularly to systems for streaming and displaying video images .

Background of the Invention:

In many situations, a scene or object is captured by multiple cameras, each of which capture a scene or object from a different angle or perspective. For example, at an athletic event multiple cameras, each at a different location, capture the action on the playing field. While each of the cameras is viewing the same event, the image available from the different cameras is different due to the fact that each camera views the event from a different angle and location. Such images can not in general be seamed into a single panoramic image.

The technology for streaming video over the Internet is well developed. Streaming video over the internet, that is, transmitting a series of images requires a substantial amount of bandwidth. Transmitting multiple streams of images (e.g. images from multiple separate cameras) or transmitting a stream of panoramic images requires an exceptionally large amount of bandwidth.

A common practice in situations where an event such as a sporting event is captured with multiple cameras, is to utilize an editor or technician in a control room to select the best view at each instant. This single view is transmitted and presented to users that are observing the event on a single screen. There are also a number of known techniques for presenting multiple views on a single screen. In one known technique, multiple images are combined into a single combined image which is transmitted and presented to users as a single combined image. With another technique the streams from the different cameras remain distinct and multiple streams are transmitted to a user who then selects the desired stream for viewing. Each of the techniques which

stream multiple images require a relatively large amount of bandwidth. The present invention is directed to making multiple streams available to a user without using an undue amount of bandwidth.

Summary of the Invention:

The present invention provides a system for capturing multiple images from multiple cameras and selectively presenting desired views to a user. Multiple streams of data are streamed to a user's terminal. One data stream (called a thumbnail stream) is used to tell the user what image streams are available. In this stream, each image is transmitted as a low resolution thumbnail. One thumbnail is transmitted for each camera and the thumbnails are presented as small images on the users screen. The thumbnail stream uses a relatively small amount of bandwidth. Another data stream (called the focus stream) contains a series of high resolution images from a selected camera. The images transmitted in this streams are displayed in a relatively large area on the viewer's screen. A user can switch the focus stream to contain images from any particular camera by clicking on the associated thumbnail. In an alternate embodiment in addition to the thumbnails from individual cameras a user is also provided with a thumbnail of panoramic image (e. g. a full 360 degree panorama or a portion thereof) which combines into a single image, the images for multiple cameras. By clicking at a position on the panoramic thumbnail, the focus stream is switched to an image from viewpoint or view window located at the point in the panorama where the user clicked. In other alternate embodiments a variety of other data streams are also sent to the user. The other data streams sent to the user can contain (a) audio data, (b) interactivity markup data which describes regions of the image which provide interactivity opportunities such as hotspots, (c) presentation markup data which defines how data is presented on the user's screen, (d) a telemetry data stream which can be used for various statistical data. In still another embodiment one data stream contains a low quality base image for each data stream. The base images serve as the thumbnail images. A second data

stream contains data that is added to a particular base stream to increase the quality of this particular stream and to create the focus stream.

Brief Description of Drawings:

Figure 1 is an overall high level diagram of a first embodiment of the invention.
Figure 2 illustrates the view on a user's display screen.
Figure 3 is a block diagram of a first embodiment of the invention.
Figure 3A illustrates how the thumbnail data stream is constructed.
Figure 4A illustrates how the user interacts with the system.
Figures 4B to 4F show in more detail elements shown in Figure 4A.
Figure 5 illustrates how clips are selected.
Figure 6 is an overview of the production process.
Figure 7 is a system overview diagram.
Figure 8 illustrates the clip production process
Figure 9 illustrates the display on a user's display with an alternate embodiment of the invention.
Figure 10 illustrates an embodiment of the invention which includes additional data streams.
Figures 11 and 11A illustrate an embodiment of the invention where the thumbnail images are transmitted and displayed with the focus view.
Figure 12 illustrates the interaction between the client and the server over time.

Detailed Description:

An overall diagram of a first relatively simplified embodiment of the invention is shown in Figure 1. In the first embodiment of the invention, an event 100 is viewed and recorded by the four cameras 102A to 102D. The event 100 may for example be a baseball game. The images from cameras 102A to 102D is captured and edited by system 110. System 110 creates two streams of video data. One stream is the images captured by "one" selected camera. The second stream consists of "thumbnails" (i.e. small low resolution images) of the images captured by each of the four cameras 102A to 102D.

The two video streams are sent to a user terminal and display 111. The images visible to the user are illustrated in Figure 2. A major portion of the display is taken by the images from one particular camera. This is termed the focus stream. On the side of the display are four thumbnail images, one of which is associated with each of the camera 102A to 102D. It is noted that the focus stream requires a substantial amount of bandwidth. The four thumbnail images have a lower resolution and all four thumbnail images can be transmitted as a single data stream. Examples of the bandwidth used by various data streams are given below.

Figure 3 illustrates a the components in a system used to practice the invention and it shows how the user interacts with the system. Camera system 300 (which includes camera 102A to 102B) provides images to unit 301 which edits the image streams and which creates the thumbnail image stream. The amount of editing depends on the application and it will be discussed in detail later. Figure 3A illustrates how the thumbnail data stream is created. The data stream from each camera and the thumbnail data stream are provided to stream control 302.

The user 306 can see a display 304. An example of what appears on display 304 is shown in Figure 2. The user has an input device (for example a mouse) and when the user "clicks on" anyone of the thumbnails, viewer software 303 sends a message to control system 302. Thereafter images from the camera associated with the thumbnail which was clicked are transmitted as the focus stream.

Figure 3A is a block diagram of the program that creates the thumbnail data stream. First as indicated by block 331, a low resolution version of each data stream is created. Low resolution images can, for example, be created by selecting and using only every fourth pixel in each image. Creating the low resolution image in effect shrinks the size of the images. As indicated by block 332, if desired the frame rate can be reduced by eliminating frames in order to further reduce the bandwidth required. The exact amount that the resolution is

reduced depends on the particular application and on the amount of bandwidth available. In general a reduction in total pixel count of at least five to one is possible and sufficient. Finally, as indicated by block 333 The corresponding thumbnail images from each data stream are placed next to each other to form composite images . The stream of these composite images is the thumbnail data stream. It should be noted that while in the data stream the thumbnails are next each other, when they are displayed on the client machine, they can be displayed in any desired location on the display screen.

The details of a first embodiment of the invention are given in Figures 4A to 4F. In this first embodiment of the invention, system 110 includes a server 401 which streams video to a web client 402 as indicated in Figure 4A. The server 401 takes the four input streams A to D from the four camera 102A to 102 D and makes two streams T and F. Stream T is a thumbnail stream, that is, a single stream of images wherein each image in the stream has a thumbnail image from each of the cameras. Stream F is the focus stream of images which transmits the high resolution images which appear on the user's display. As shown in Figure 2, the users display shows the four thumbnail images and a single focus stream.

The web client 402 includes a stream selection control 403. This may for example be a conventional mouse. When the user, clicks on one of the thumbnails, a signal is sent to the server 401 and the focus stream F is changed to the stream of images that coincides with the thumbnail that was clicked. In this embodiment server 401 corresponds to stream control 302 shown in Figure 3 and client 402 includes components 303, 304 and 305 shown in Figure 3. The details of the programs in server 401 and client 402 are shown in Figures 4B to 4E and are described later.

An optional procedure that can be employed to give a user the illusion that the change from one stream to another stream occurs instantaneously is illustrated in Figure 4F. Figure 4F shows a sequence of steps that can take place when

the user decides to change the focus stream to a different camera. It is noted that under normal operation, a system receiving streaming video buffers the data at the input of the client system to insure continuity in the event of a small delay in receiving input . This is a very common practice and it is indicated by block 461. When a command is given to change the focus stream, if the procedure shown in Figure 4F is not used, there will be a delay in that when the client begins receiving the new stream, it will not be displayed until the buffer is sufficiently filled. This delay can be eliminated using the technique illustrated in Figure 4F. With this technique when a viewer issues a command to change the focus stream the large image on the viewer's screen is immediately changed to an enlarged image from the thumbnail of the camera stream newly requested by the user. This is indicated by block 463. That is, the low resolution thumbnail from the desired camera is enlarged and used as the focus image. This insures that the focus image changes as soon as the user indicates that a change is desired. The buffer from the focus data stream is flushed and it begins filling with the images from the new focus stream as indicated by blocks 464 and 465. As indicated by block 466, when the buffer is sufficiently full of images from the new stream, the focus image is changed to a high resolution image from this buffer.

As indicated by block 301, the data streams from the cameras are edited before they are sent to users. It is during this editing step that the thumbnail images are created as indicated in Figure 3A. The data streams are also compressed during this editing step. Various known types of compression can be used.

Figure 5 illustrates another type of editing step that may be performed. The entire stream of images from all the cameras need not be streamed to the viewer. As illustrated in Figure 5, sections of the streams, called "clips" can be selected and it is these clips that are sent to a user. As illustrated in figure 5, two clips C1 and C2 are made from the video streams A to D. In general the clips would be compressed and stored on a disk file and called up when there

is a request to stream them to a user. For example, a brief description of clips showing the key plays from a sporting event can be posted on a web server, and a user can then select which clips are of interest. A selected clip would then be streamed to the user. That is, the thumbnail images and a single focus stream would be sent to a user. The streaming would begin with a default camera view as the focus view. When desired, the user can switch the focus stream to any desired camera by clicking on the appropriate thumbnail. With the first embodiment of the invention, files such as clips are stored on the server in a file with a ".pan" file type. The pan file would have the data stream from each camera and the thumbnail data stream for a particular period of time.

The first embodiment of the invention is made to operate with the commercially available streaming video technology marketed by RealNetworks Inc. located in Seattle, Washington. RealNetworks Inc. markets a line of products related to streaming video including products that can be used to produce streaming video content, products for servers to stream video over the Internet and video players that users can use to receive and watch streamed video which is streamed over the Internet. Figures 4B and 4F show the units 401 and 402 in more detail.

As indicated in Figure 4B, the web server 401 is a conventional server platform such as an Intel processor with an MS Windows NT operating system and an appropriate communications port. The system includes a conventional web server program 412. The web server program 412 can for example be the program marketed by the Microsoft Corporation as the "Microsoft Internet Information Server". A video streaming program 413 provides the facility for streaming video images. The video streaming program 413 can for example be the "RealSystem Server 8" program marketed by Real networks Inc. Programs 412 and 413 are commercially available programs. While the programs 412 and 413 are shown resident on a single server platform, these two programs could be on different server platforms. Other programs from

other companies can be substituted for the specific examples given. For example the Microsoft corporation markets a streaming server termed the "Microsoft Streaming Server" and the Apple Corporation markets streaming severs called QuickTime and Darwin.

In the specific embodiment shown "video clips" are stored on a disk storage sub-system 411. Each video clip has a file type ".pan" and it contains the video streams from each of the four cameras and the thumbnail stream. When system receives a URL calling for one of these clips, the fact that the clip has a file type ".pan" indicates that the file should be processed by plug in 414.

One of the streams stored in a pan file is a default stream and this stream is sent as the focus stream until the user indicates that another stream should be the focus stream. Plug in 414 process requests from the user and provides the appropriate T and F streams to streaming server 413 which sends the streams to the user. The components of the plug 414 are explained later with reference to figure 4D.

As illustrated in Figure 4C, client 402 is a conventional personal computer with a number of programs. The client 402 includes a Microsoft Windows operating system 422, and a browser program 423. The browser 423 can for example be the Microsoft Internet Explorer browser. Streaming video is handled by a commercially available program marketed under the name: "RealPlayer 8 Plus" by RealNetworks Inc. Programs 422, 423 and 424 are conventional commercially available programs. Other similar programs can also be used. For example Microsoft and Apple provide players for streaming video. A plug in 425 for the Real Player 424 renders images from pan files, that is, plug in 425 handles the thumbnail and focus data streams and handles the interaction between the client 402 and the plug in 414 in the server 401. The components in plug in 425 are given in Figure 4E.

Figures 4D and 4E are block diagrams of the programming plug in 414 and 425. Plug in 414 is shown in Figure 4D. When the server encounter a request to stream a file with the file type ".pan", it retrieves this file from disk storage subsystem 411 (unless the file is made available to the server via some other input). The file is then transferred to plug in 414. This is indicated by block 432. Commands from the user i.e. "clicks" on a thumbnail, or other types of input from the user when a pan file is being streamed are also sent to this plug in 414. As indicated by block 435, plug in 435 selects the thumbnail stream and either a default or a requested stream from the pan file. As indicated by block 437, the thumbnail stream and the selected focus stream are sent to the "Real System Server 8" program. In alternate embodiments, other streams are also available in pan files. These other streams are selected and sent to the "Real System Server 8" program as appropriate in the particular embodiment.

Figure 4E is a block diagram of the programming components in the plug in 425 on the client machine. When the Real Player 8 Plus 424 encounters data from a pan files, the data is sent to plug in 425. Figure 4E shows this data as block 451. The stream manager recognizes the different types of data streams and sends the data to an appropriate handler 454A to 454C. Data may be temporarily stored in a cache and hence, as appropriate the data handler retrieves data from the cache. Each handler is specialized and can handle a specific type of stream. For example one handler handles the thumbnail stream and another handler handles the focus stream. The thumbnail handler divides the composite images in the thumbnail stream into individual images. The handlers use a set of decoding, decompression and parsing programs 455A to 455B as appropriate. The system may include more handlers than shown in the figure if there are more kinds of data streams. Likewise the system may include as many decoder, decompression and parsing programs as required for the different types of streams in a particular embodiment . The brackets between the handlers and the decoders in Figure 4E indicate that any handler can use any appropriate decoder and parser to process image data as appropriate. The decompressed and parsed data is sent to a rendering

program 456 which sends the data to the real play input port to be displayed. A controller 443 controls gating and timing of the various operations.

It should be clearly noted the specific examples given in Figures 4A to 4E are merely examples of a first simplified embodiment of the invention. For example instead of working with a web server, the invention could work with other types of servers such as an intranet server or a streaming media server or in fact the entire system could be on a single computer with the source material being stored on the computer's hard disk. The interaction between the sever 401 and the client 402, and the manner the server responds to the client 402 is explained in detail later with reference to Figure 12. It should be noted that all of the components shown in Figures 4A to 4E (other than the server platform and personal computer) are software components.

Figure 6 illustrates the system in a typical setup at a sporting event. The cameras and the sporting event are in stadium 601. The output from the camera goes to a video production truck 602 which is typical owned by a TV network. Such trucks have patch panels at which the output from the cameras can be made available to equipment in a clips production truck 603. The clip production truck 603 generates the clips and sends them to a web site 604.

Figure 7 is a system overview of this alternate embodiment. The "feed" from stadium cameras 701 goes to patch panel 702 and then to a capture station 703. At station 703 operator 1 makes the clip selections as illustrated in Figure 5. He does this by watching one of the channels and when he sees interesting action he begins capturing the images from each of the camera. The images are recorded digitally. The images can be digitally recorded with commercially available equipment. Cutting clips from the recorded images can also be done with commercially available equipment such as the "Profile™" and "Kalypso™" Video Production family of equipment marketed by Grass Valley Group Inc. whose headquarters are in Nevada City, California.

As shown in Figure 8 when a clip is selected as indicated at 801, the clip is stored and it is given a name as indicated on display 703. The stored clips are available to the operator of the edit station 704. At the edit station, the clip can be edited, hot spots can be added and voice can be added. Hot spots are an overlay provided on the images such that if the user clicks at a particular position on an image as it is being viewed, some action will be taken. Use of hot spots is a known technology. When the editing is complete the clips are compressed and posted on web site 705.

Figure 9 illustrates what a user sees with another alternate embodiment of the invention. The alternative embodiment illustrated in Figure 9 is designed for use with multiple cameras which record images which can be seamed into a panorama. Cameras which record multiple images which can be seamed into a panorama are well known. For example see co-pending application serial number 09/338,790, filed 6/23/99 and entitled "A System for Digitally Capturing and Recording Panoramic Movies".

The embodiment shown in Figure 9 is for use with a system that captures six images such as the camera shown in the referenced co-pending application (which is hereby incorporated herein by reference). The six images captured by the camera are: a top, a bottom, a left side, a right side, a front and a back images (i.e. there is a lens on each side of a cube). These images can be seamed into a panorama in accordance with the prior art and stored in a format such as an equi-rectangular or cubic format. With this alternative embodiment, the user sees a display such as that illustrated in Figure 9. At the top center of the display is a thumbnail 901 of a panorama. The panoramic image is formed by seaming s together into one panoramic image, the individual images from the six cameras. Six thumbnails of images from the cameras (the top, bottom, left side, right side, front and back of the cube) are shown along the right and left edges of the display. If a user clicks on any one of the six thumbnails, on the right and left of the screen, the focus stream switched to that image stream as in the first embodiment. It is noted that with a panoramic image, it is usual

for a viewer to select a view window and then see the particular part of the panorama which is in the selected view window. If the user clicks anywhere in the panorama 901, the focus stream is changed to a view window into the panorama which is centered at the point where the user clicked. With this embodiment, stream control has as one input a panoramic image and the stream control selects a view window from the panorama which is dependent upon where the user clicks on the thumbnail of the panorama. The image from this view window is then streamed to the user as the focus image.

In other alternative embodiments which show a thumbnail of a panorama, as described above, in addition to (or in place of) the thumbnails of the individual camera views from the camera which were used to record the panorama, thumbnails from other camera are provided. These additional cameras may be cameras which are also viewing the same event, but from a different vantage point. Alternatively they can be from some related event.

A somewhat more complicated alternate embodiment of the invention is shown in Figure 10. In the embodiment illustrated in Figure 10, a server 910 receives eight streams S1 to S8. The eight streams include four streams S5 to S8 that are similar to the video streams described with reference to the previously described embodiment. These four streams include a stream S8 where each image contains a thumbnail of the other images and three video streams designated V1 to V3.

The server selects the streams that are to be streamed to the user as described with the first embodiment of the invention. The selected streams are then sent over a network (for example over the Internet) to the client system.

The additional data streams provided by this embodiment of the invention include an audio stream S4, an interactivity markup stream S3, a presentation markup stream S2 and a telemetry data stream S1. The audio stream S4 provides audio to accompany the video stream. Typically there would be an

single audio stream which would be played when any of the video streams are viewed. For example, there may be a play by play description of a sporting event which would be applicable irrespective of which camera is providing the focus stream. However, there could be an audio stream peculiar to each video stream.

The interactivity markup stream S3 describes regions of the presentation which provide for additional user interaction. For example there may be a button and clicking on this button might cause something to happen. The interactivity markup stream consists of a series of encoded commands which give type and position information. The commands can be in a descriptive language such as XML encoded commands or commands encoded in some other language. Such command languages are known and the ability to interpret commands such as XML encoded commands is known.

The presentation markup stream provides an arbitrary collection of time synchronized images and data. For example, the presentation markup stream can provide a background image for the display and provide commands to change this background at particular times. The presentation mark up stream may provide data that is static or dynamic. The commands can, for example, be in the form of XLM encoded commands.

The telemetry data stream S1 can provide any type of statistical data. For example this stream can provide stock quotes or player statistics during a sporting event. Alternatively the stream could provide GPS codes indicating camera position or it could be video time codes.

Yet another alternate embodiment of the invention is shown in Figure 11. With the embodiment shown in Figure 11, there is not a separate video stream for the thumbnail images. In this embodiment, instead of having a separate stream for the thumbnail, the thumbnails are transmitted as part of the video streams V1, V2 and V3. A set of the thumbnails is included in each of the

video streams. Hence, irrespective of which video stream is selected for the focus steam, the user will have available thumbnails of the other streams. Figure 11A illustrates the display showing an image from the focus stream with the thumbnails on the bottom as part of this image.

A key consideration relative to video streaming is the bandwidth required. If unlimited bandwidth were available, all the data streams would be sent to the client. The present invention provides a mechanism whereby a large amount of data, for example data from a plurality of camera, can be presented to a user over a limited bandwidth in a manner such that the user can take advantage of the data in all the data streams. The specific embodiments shown relate to data from multiple camera that are viewing a particular event. However, the multiple streams need not be from cameras. The invention can be used in any situation where there are multiple streams of data which a user is interested in monitoring via thumbnail images. With the invention, the user can monitor the multiple streams via the thumbnail images and then make any particular stream the focus stream which becomes visible in an high quality image. Depending upon the amount of bandwidth available there could be a large number of thumbnails and there may be more than one focus stream that is sent and shown with a higher quality image.

The flowing table shows the bandwidth requirements of various configurations.

Main Video Size 320x240

Number Video Streams	2	2	3	3	4	4
Video Stream Vertical	240	240	240	240	240	240
Video Stream Horizontal	320	320	320	320	320	320
Thumbnail Vertical	100	100	100	100	100	100
Thumbnail Horizontal	75	75	75	75	75	75
Video frame rate	7	15	7	15	7	15
Color Depth (bits)	24	24	24	24	24	24
MPEG4 Video Compression ratio	150	150	150	150	150	150
Presentation Video Bandwidth	**188832**	**404640**	**283248**	**606960**	**377664**	**809280**
Shaped Video Bandwidth	**102816**	**220320**	**111216**	**238320**	**119616**	**256320**
Number Audio Streams	1	1	1	1	1	1

Audio bitrate	30000	30000	30000	30000	30000	30000
Presentation Audio Bandwidth	**30000**	**30000**	**30000**	**30000**	**30000**	**30000**
Number Telemetry Streams	1	1	1	1	1	1
Telemetry bit rate	500	500	500	500	500	500
Presentation Telemetry Bandwidth	**500**	**500**	**500**	**500**	**500**	**500**
Number Presentation Markup Stream	1	1	1	1	1	1
Markup bitrate	2500	2500	2500	2500	2500	2500
Presentation Markup Bandwidth	**2500**	**2500**	**2500**	**2500**	**2500**	**2500**
Number Interactivity Markup Stream	1	1	1	1	1	1
Markup bitrate	1000	1000	1000	1000	1000	1000
Presentation Markup Bandwidth	**1000**	**1000**	**1000**	**1000**	**1000**	**1000**
Presentation Bandwidth (bps)	**222832**	**438640**	**317248**	**640960**	**411664**	**843280**
Presentation Bandwidth (Kbs)	**217.61**	**428.36**	**309.81**	**625.94**	**402.02**	**823.52**
Presentation Bandwidth (KBs)	**27.20**	**53.54**	**38.73**	**78.24**	**50.25**	**102.94**
Shaped Bandwidth	**136816**	**254320**	**145216**	**272320**	**153616**	**290320**
Shaped Streaming (Kbs)	**133.61**	**248.36**	**141.81**	**265.94**	**150.02**	**283.52**
Shaped Streaming (KBs)	**16.70**	**31.04**	**17.73**	**33.24**	**18.75**	**35.44**

The interaction between the server and the client is illustrated in Figure 12. Figure 12 illustrates the three components of the system. The components are:

The client: The client is operated by a user. It displays the presentation content received from the server. It instructs the server to change Focus streams, play forward, backwards, fast forward, fast reverse, replay pause and stop.

The server: The server responds to client requests. The presentation source: The presentation source could be disk storage, a remote server, or a feed from a computer that is generating a presentation from live inputs.

As illustrated in Figure 12, the process begins when the client requests a presentation as indicated by arrow 991. This creates a server session and the server begins accessing the presentation from the presentation source and providing it to the server as indicated by arrow 992. The server then being streaming this information to the client. At this point the focus stream is a default stream. The client's screen is configured according to the layout

information given in the presentation mark up stream. For example this could be XML encoded description commands in the presentation markup stream. In the example given, at this point the client requests that the focus stream change. This is sent to the server as indicated by arrow 994.

When the server receives the command, it stops streaming the old focus stream and starts streaming the new focus stream as indicated by arrow 995. A new layout for the user's display is also sent as indicated by arrow 996. It is noted that a wide variety of circumstances could cause the server to send to the client a new layout for the users display screen. When the client receives the new display layout, the display is reconfigured.

Arrow 997 indicates that the user can request an end to the streaming operation. Upon receipt of such a request or when the presentation (e.g. the clip) ends, the server stops the streaming operation and ends access to the presentation source as indicated by arrows 998. The server also ends the connection to the client as indicated by arrow 999 and the server session ends. It should be understood that the above example is merely illustrative and a wide variety of different sequences can occur.

Another embodiment of the invention operates by sending base information to create the thumbnail images and additional information to create the focus image. The user sees the same display with this embodiment as the user sees with the previously described embodiments; however, this embodiment uses less bandwidth. With this embodiment, the focus data stream is not a stream of complete images. Instead, the focus stream is merely additional information, that can be added to the information in one of the thumbnails images to create a high resolution image. The thumbnail images provide basic information which creates a low resolution thumbnail. The focus stream provides additional information which can be added to the information in a thumbnail to create a high resolution large image.

1 The following table illustrates the bandwidth savings:

Main Video Size 320x240	Previously embodiment	Using Base and Enhancement Layers
Number of Input Video Streams	3	3
Number Base Layer Streams	0	3
Number Enhancement Layer Streams	0	3
Video Stream Vertical	240	240
Video Stream Horizontal	320	320
Thumbnail Vertical	75	75
Thumbnail Horizontal	100	100
Video frame rate	15	15
Color Depth (bits)	24	24
MPEG4 Video Compression ratio	150	150
Presentation Video Bandwidth	**606960**	**552960**
Shaped Video Bandwidth	**238320**	**184320**
Number Audio Streams	1	1
Audio bitrate	30000	30000
Presentation Audio Bandwidth	**30000**	**30000**
Number Telemetry Streams	1	1
Telemetry bit rate	500	500
Presentation Telemetry Bandwidth	**500**	**500**
Number Presentation Markup Stream	1	1
Markup bitrate	2500	2500
Presentation Markup Bandwidth	**2500**	**2500**
Number Interactivity Markup Stream	1	1
Markup bitrate	1000	1000
Presentation Markup Bandwidth	**1000**	**1000**
Presentation Bandwidth (bps)	**640960**	**586960**
Presentation Bandwidth (Kbs)	**625.94**	**573.20**
Presentation Bandwidth (KBs)	**78.24**	**71.65**
Shaped Bandwidth	**272320**	**218320**
Shaped Streaming (Kbs)	**265.94**	**213.20**
Shaped Streaming (KBs)	**33.24**	**26.65**

2

3 Subdividing the image data can further reduce bandwidth by allowing optimized
4 compression techniques to be used on each subdivision. Subdivisions may be
5 made by any desirable feature of the imagery, such as pixel regions,

foreground/background, frame rate, color depth, resolution, detail type, etc., or any combination of these. Each data stream can be compressed using a technique that preserves the highest quality for a given bandwidth given its data characteristics. The result is a collection of optimally compressed data streams, each containing a component of the resultant images. With this embodiment, each thumbnail image stream is constructed on the client by combining several of these data streams, and its corresponding focus image stream is constructed on the client by combining the thumbnail streams (or thumbnail images themselves) and more data streams.

For example, consider a multiple view video that consists of different views of live action characters superimposed against the same static background image. The client sees a low-resolution thumbnail stream for each view and a high-resolution focus stream of one of them. These view streams could be compressed as described before, with a low-resolution thumbnail stream and additional data streams for turning them into high-resolution focus streams. However, additional bandwidth savings can be realized if two features of the images streams are utilized: a) the frame rate of the background image is different than the foreground, specifically, the background image is static throughout the entire presentation, so only one image of it ever needs to be sent regardless of how many image frames the presentation is, and b) the same background image is used for all the view streams, so only one copy of the background image needs to be sent and can be reused by all the view streams. In order to realize this bandwidth savings, a foreground/background subdivision may be made to the video data in the following way:

a) A data stream containing a single low-resolution background image that is reused to generate all the thumbnail images

b) Data streams containing low-resolution foreground images for the thumbnail views, one stream per view.

c) A data stream containing additional data to boost the low-resolution background image to become the high-resolution background image.

d) Data streams containing additional data for boosting the low-resolution foreground images to become high-resolution foreground images.

In this embodiment, each image in the thumbnail stream is generated on the client by combining the low-resolution background image with the appropriate low-resolution foreground image. Each image in the focus stream is generated on the client by: adding the additional background image data to the low-resolution background image to generate the high-resolution foreground image, adding the additional foreground image data to the low-resolution foreground image to generate the high-resolution foreground image, and then combining the high-resolution foreground and background images to generate the final focus-stream image.

As another example, consider a video where each stream contains a view of a subject against a blurry background, such as one might see at a sporting event where a cameraman has purposely selected camera settings that allow the player to be in crisp focus while the crowd behind the player is significantly blurred. The client sees a low-resolution thumbnail stream for each view and a high-resolution focus stream of one of them. These views could be compressed with a quality setting chosen to preserve the detail in the player. However, bandwidth savings could be realized by utilizing the fact that the blurry crowd behind the player is unimportant to the viewer and can therefore be of lower quality. In order to realize this bandwidth savings, a pixel region subdivision can be made to the image data in the following way:

a) A data stream containing the player region in low resolution, for the thumbnail images.

b) A data stream containing the remaining image region in low-resolution, for the thumbnail images. This image region would be compressed with a lower quality than that used for the player region.

c) An additional data stream, one per focus view, for boosting the low-resolution player region into a high-resolution player region.

d) An additional data stream, on per focus view, for boosting the remaining image region from low-resolution to high-resolution. This image region would be compressed with a lower quality than that used for the player region.

Each image in the thumbnail stream is generated on the client by combining the player region with the rest of that image. Each image in the focus stream is generated on the client by: adding the additional player region data to the low-resolution player image to generate the high-resolution player image, adding the additional remaining image data to the low-resolution remaining image region generate the high-resolution remaining image region, and then combining the two regions to generate the final focus-stream image.

As another example, consider a video where each stream contains fast-moving objects that are superimposed on slowly changing backgrounds. The client sees a low-resolution thumbnail stream for each view and a high-resolution focus stream of one of them. Each stream of video could use a frame rate that allows the fast-moving object to be displayed smoothly. However, bandwidth savings could be realized by utilizing the fact that the slowly changing background differs little from one frame to the next, while the fast-moving object differs significantly from one frame to the next. In order to realize this bandwidth savings, a pixel region subdivision must be made to the image data in the following way:

a) A data stream containing the fast-moving object regions in low resolution, for the thumbnail images. This stream uses a fast frame rate.

b) A data stream containing the remaining image region in low-resolution, for the thumbnail images. This stream uses a slower frame rate than what was used for the fast-moving object region.

c) An additional data stream, one per focus view, for boosting the low-resolution fast-moving object region into a high-resolution fast-moving object region. This stream uses a fast frame rate.

d) An additional data stream, on per focus view, for boosting the remaining image region from low-resolution to high-resolution. This stream uses a slower frame rate than what was used for the fast-moving object region.

In this embodiment, each image in the thumbnail stream is generated on the client by combining the fast-moving object region with the most-recent frame of the rest of that image. Each image in the focus stream is generated on the client by: adding the additional fast-moving object region data to the low-resolution fast-moving object image to generate the high-resolution fast-moving object image, adding the additional remaining image data to the low-resolution remaining image region to generate the high-resolution remaining image region, and then combining the high-resolution fast-moving object regions with the most recent frame of the remaining image region to generate the final focus-stream image.

As another example, consider a video where each stream contains well-lit subjects in front of a differently lit background that results in a background that is shades of orange. The client sees a low-resolution thumbnail stream for each view and a high-resolution focus stream of one of them. Each stream of video could use the whole images as is. However, bandwidth savings could be realized by utilizing the fact that the background uses a restricted palette of orange and black hues. In order to realize this bandwidth savings, a pixel region subdivision must be made to the image data in the following way:

a) A data stream containing the image region that the well-lit subject occupies, for the thumbnail images. Full color data is retained for these images.

b) A data stream containing the remaining image region in low-resolution, for the thumbnail images. For these images, the full color data is discarded and only the brightness value part of the color data is retained, allowing fewer bits of data to be used for these images. Upon decompression, these brightness values will be used to select the appropriate brightness of orange coloration for that part of the image.

c) An additional data stream, one per focus view, for boosting the low-resolution image of the well-lit subject into a high-resolution image of the well-lit subject. Full color data is retained for this additional data.
d) An additional data stream, on per focus view, for boosting the remaining image region from low-resolution to high-resolution. For this additional data, the full color data is discarded and only the brightness value part of the color data is retained, allowing fewer bits of data to be used. Upon decompression, these brightness values will be used to select the appropriate brightness of orange coloration for that part of the image.

In this embodiment, each image in the thumbnail stream is generated on the client by combining the well-lit subject object region with the remaining image region in which the brightness values in the image were used to select the correct brightness of orange color for those parts of the image. Each image in the focus stream is generated on the client by: adding the additional well-lit subject region data to the low-resolution well-lit subject image to generate the high-resolution well-lit subject image, adding the additional remaining image data to the low-resolution remaining image region to generate the high-resolution remaining image region and using the brightness values in the image to select the correct brightness of orange color for those parts of the image, and then combining the high-resolution well-lit subject regions with the remaining image region generated earlier.

While the invention has been shown and described with respect to a plurality of preferred embodiments, it will be appreciated by those skilled in the art, that various changes in form and detail may be made without departing from the spirit and scope of the invention. The scope of applicant's invention is limed only by the appended claims.

I claim:

1) A system for capturing and displaying images comprising,
a plurality of video cameras viewing an event,
digital storage for storing the outputs from said cameras,
a first edit station with access to said stored video for selecting clips from said video streams,
a second edit station for editing the output of said first edit station,
a web site for storing the edited clips,
a user browser,
a packaging program for creating two video streams one of which is a focus stream and one of which contains a thumbnail of the images from each camera,
a control device at said browser whereby a user can signal to said web site as to which video stream should be the focus stream.

2) A system for displaying to a user a selected one of a plurality of video streams, said selected video stream being a focus stream, said system comprising,
a client system which can display said selected video stream, and a composite video containing a thumbnail image of each of said plurality of video streams,
a server which receives a plurality of video streams, and said composite video stream, and which provides a selected one of said video streams and said composite video stream to said client system, and
an input device connected to said client system whereby a user can select one of said thumbnails thereby sending a signal to said server indicating which of said plurality of video streams should be sent to said client system.

3) The system recited in claim 2 wherein said server also sends a presentation markup stream to said client machine to control the presentation of images by said client machine.

4) The system recited in claim 2 wherein said server also sends an audio stream to said client machine.

5) The system recited in claim 2 wherein said server also sends an interactivity mark up stream to said client system to describe regions of the presentation that provide additional user interaction with said system.

6) The system recited in claim 2 wherein said server also sends a stream of telemetry data to said client machine.

7) The system recited in claim 2 wherein said server also sends an audio stream, a presentation mark up stream, and an interactive markup stream to said client machine.

8) A method of streaming selected data from a plurality of cameras to a user who is viewing a display on a client machine comprising the steps of:
streaming to said client machine a focus stream containing the images from a particular one of said cameras and a second video stream, each image in which contains a thumbnail of the images from each of said cameras, and responding when a user selects a thumbnail of the images from a selected camera by making said focus stream the images from said selected camera.

9) The method recited in claim 8 wherein a presentation mark up stream is sent to said client machine to indicate to control the display of images on said client machine.

10) The method recited in claim 8 wherein an audio stream is also sent to said client machine.

11) The method recited in claim 8 wherein an interactivity mark up stream is sent to said client system to describe regions of the presentation that provide additional user interaction with said system.

12) The method recited in claim 8 wherein a stream of telemetry data is also sent to said client machine

13) A system for displaying video images comprising,
a server which has available a plurality of data streams,
a client,
first means for streaming a first video stream from said server to said client and for simultaneously streaming a second video stream from said server to said client, said second stream consisting of composite images each of which includes a plurality of low resolution images,
means at said client for receiving said streams and for displaying to a user a high resolution image and a plurality of thumbnails which indicate other streams which are available at said server,
means for allowing said user to indicate which of the streams indicated by said thumbnails, said user would like to form said focus stream.

14) The system recited in claim 13 wherein said server and said client are located on one physical computer system.

15) The system recited in claim 2 wherein said server and said client are located on one physical computer system.

16) The system recited in claim 13 wherein said first stream contains base data which can form low resolution thumbnail images and said second stream contains enhancement data which can be added to a low resolution image to form an enhanced image.

17) The system recited in claim 13 wherein one of said low resolution images is an panoramic image and wherein said high resolution image can be a view window from said panorama.

18) The system recited in claim 2 wherein both of said composite images and said high resolution images are contained in the same data stream.

19) The system recited in claim 13 wherein a background image is steamed to said client and said high resolution image and said thumbnail images are superimposed over said background image.

20) The system recited in claim 13 wherein said high resolution images presented to said user include hot spots which can be used to activate commands.

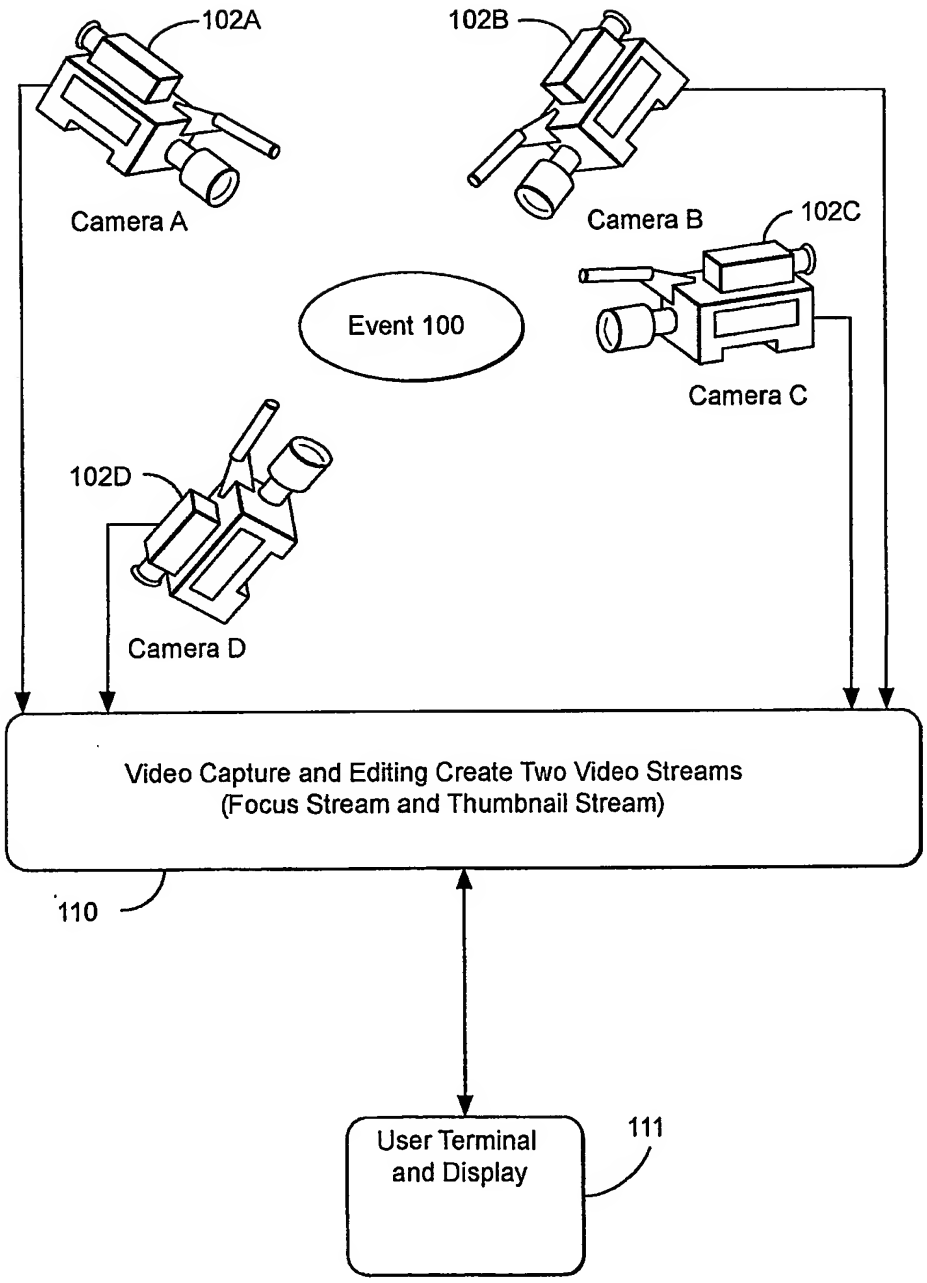

102A
102B
Camera A
Camera B
102C
Event 100
Camera C
102D
Camera D
Video Capture and Editing Create Two Video Streams
(Focus Stream and Thumbnail Stream)
110
User Terminal
and Display
111

FIG. 1

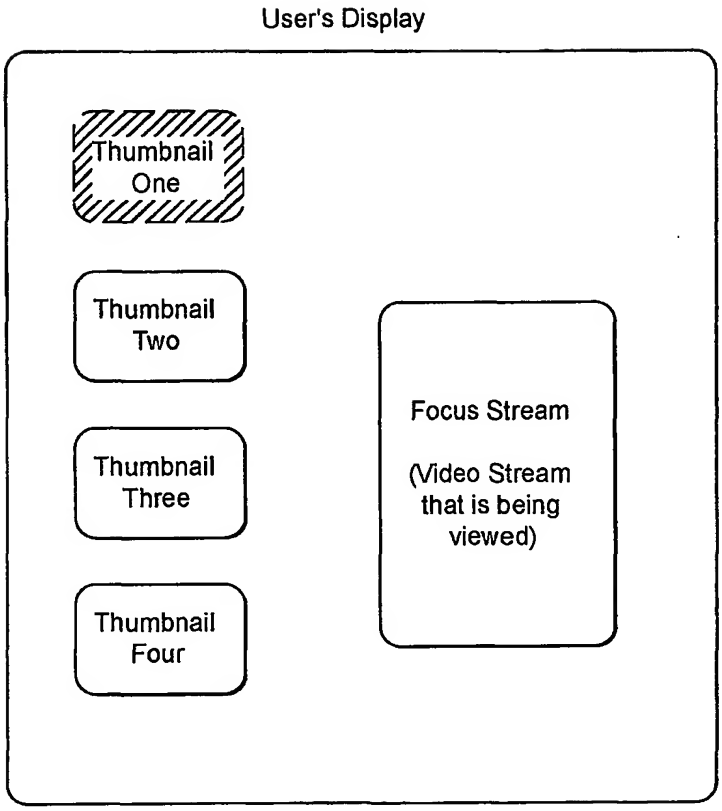

User's Display
Thumbnail One
Thumbnail Two
Thumbnail Three
Thumbnail Four
Focus Stream
(Video Stream that is being viewed)

FIG. 2

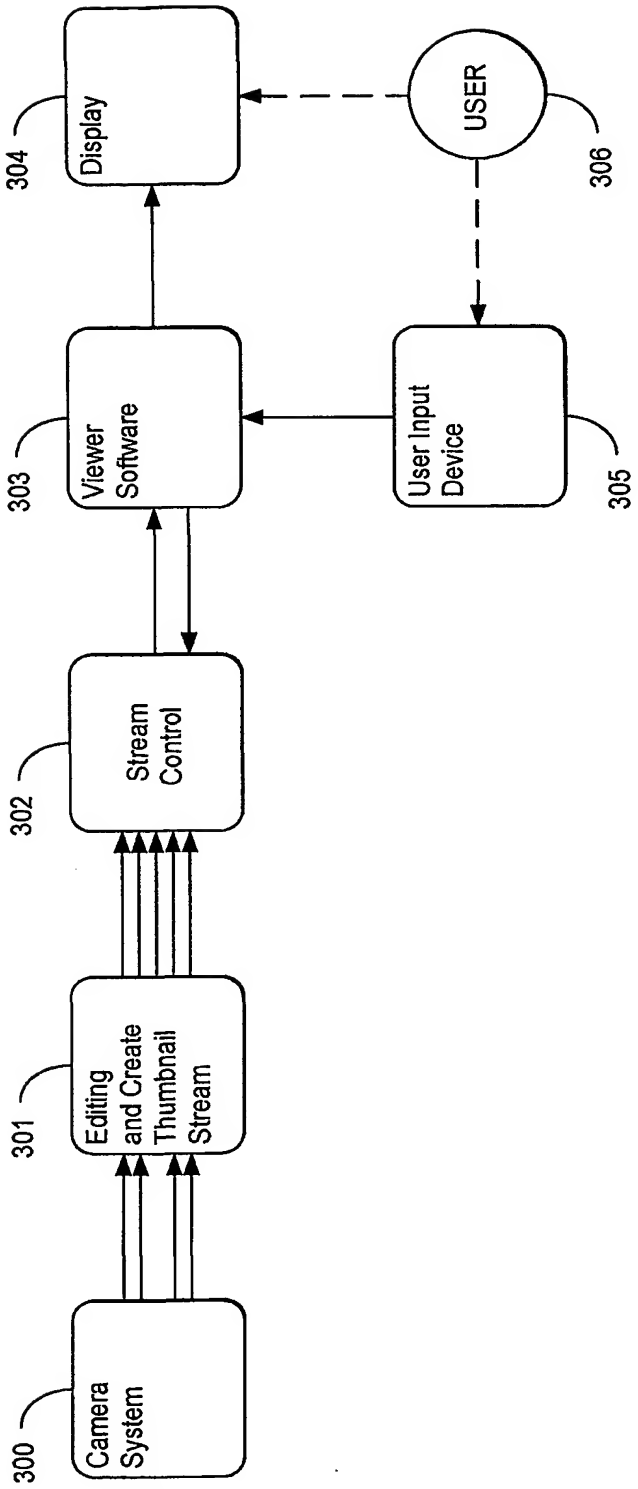

Display
304
USER
306
Viewer Software
303
User Input Device
305
Stream Control
302
Editing and Create Thumbnail Stream
301
Camera System
300

FIG. 3

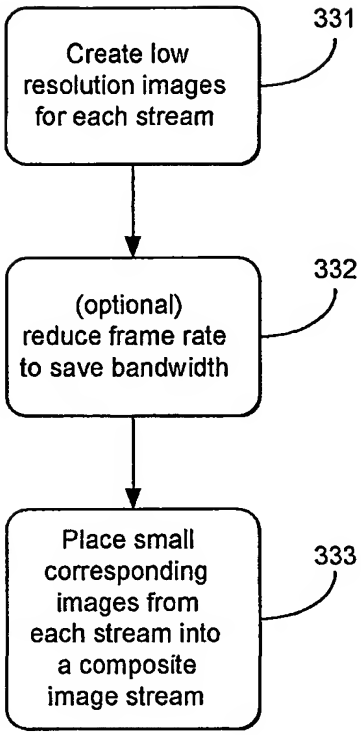

Create low resolution images for each stream
331
(optional) reduce frame rate to save bandwidth
332
Place small corresponding images from each stream into a composite image stream
333

FIG. 3A

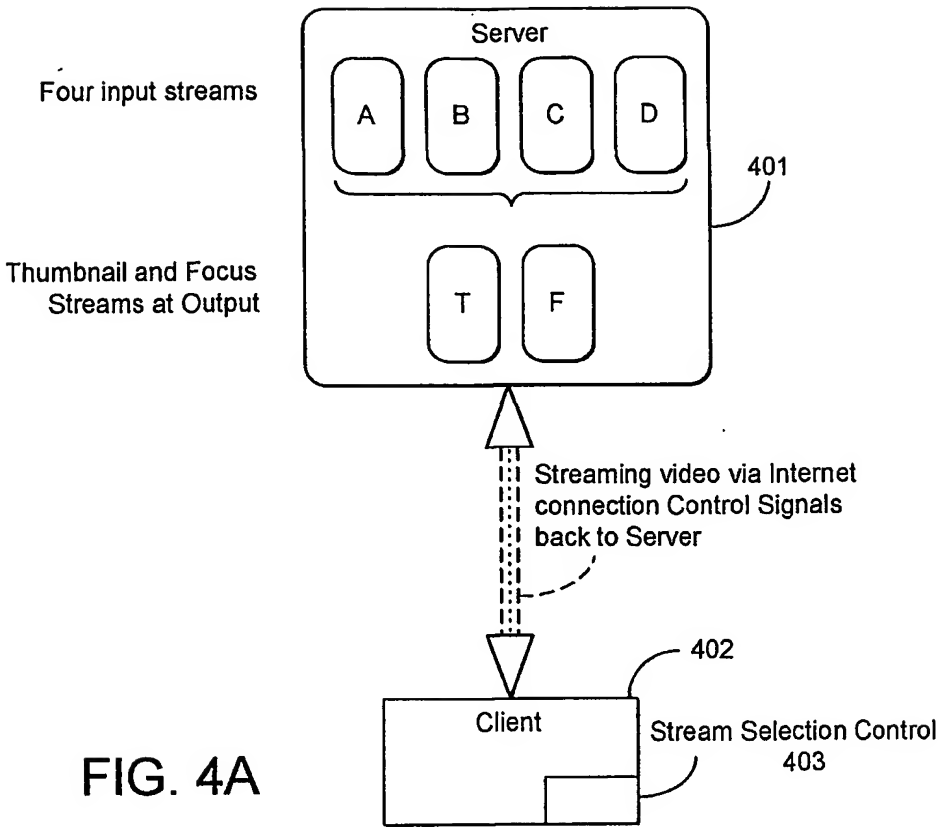


FIG. 4A

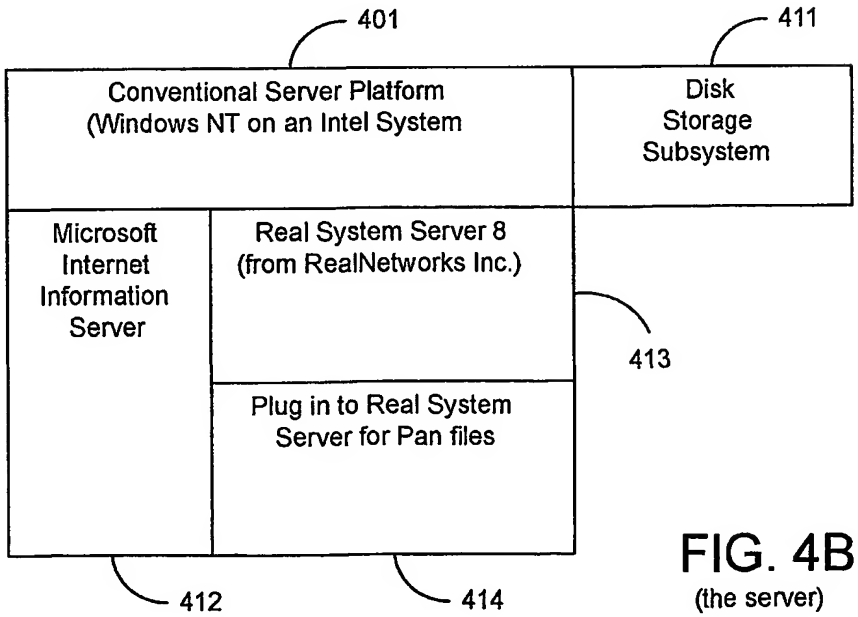


FIG. 4B
(the server)

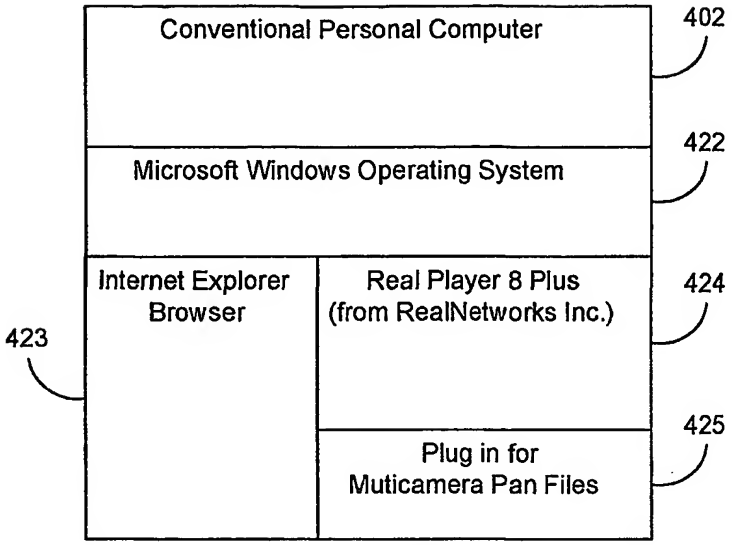

Conventional Personal Computer
402
Microsoft Windows Operating System
422
Internet Explorer Browser
Real Player 8 Plus (from RealNetworks Inc.)
424
423
Plug in for Muticamera Pan Files
425
(the client)

FIG. 4C

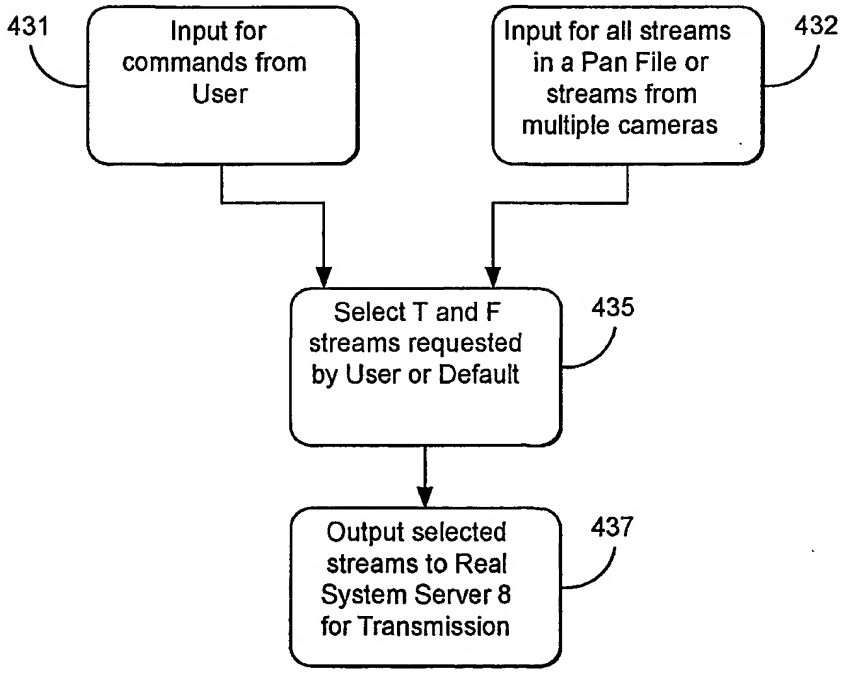

431
Input for commands from User
Input for all streams in a Pan File or streams from multiple cameras
432
Select T and F streams requested by User or Default
435
Output selected streams to Real System Server 8 for Transmission
437
(server plug in for pan files)

FIG. 4D

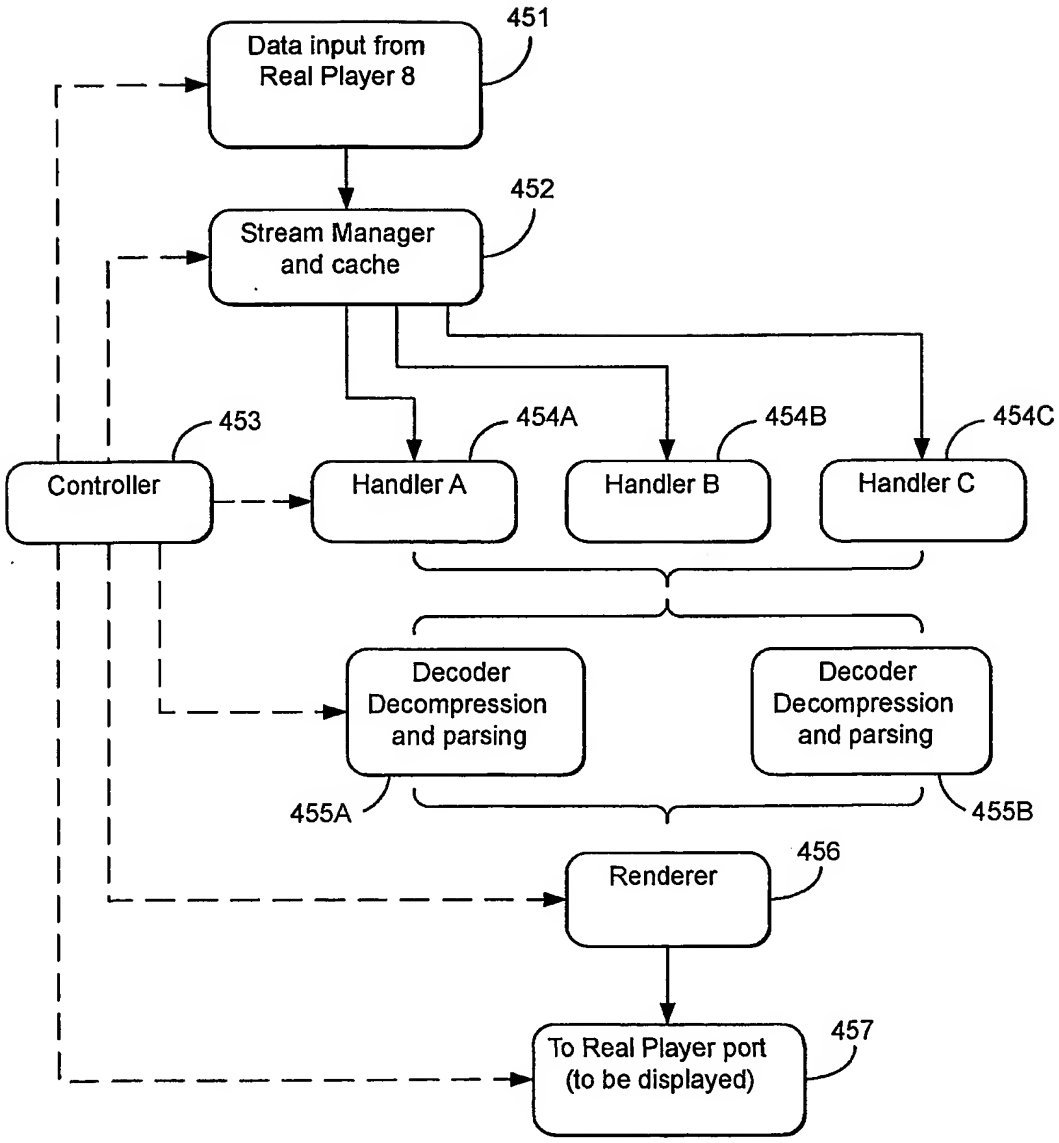

Data input from Real Player 8
451
Stream Manager and cache
452
Controller
453
Handler A
454A
Handler B
454B
Handler C
454C
Decoder Decompression and parsing
Decoder Decompression and parsing
455A
455B
Renderer
456
To Real Player port (to be displayed)
457
(client plug in for pan files)

FIG. 4E

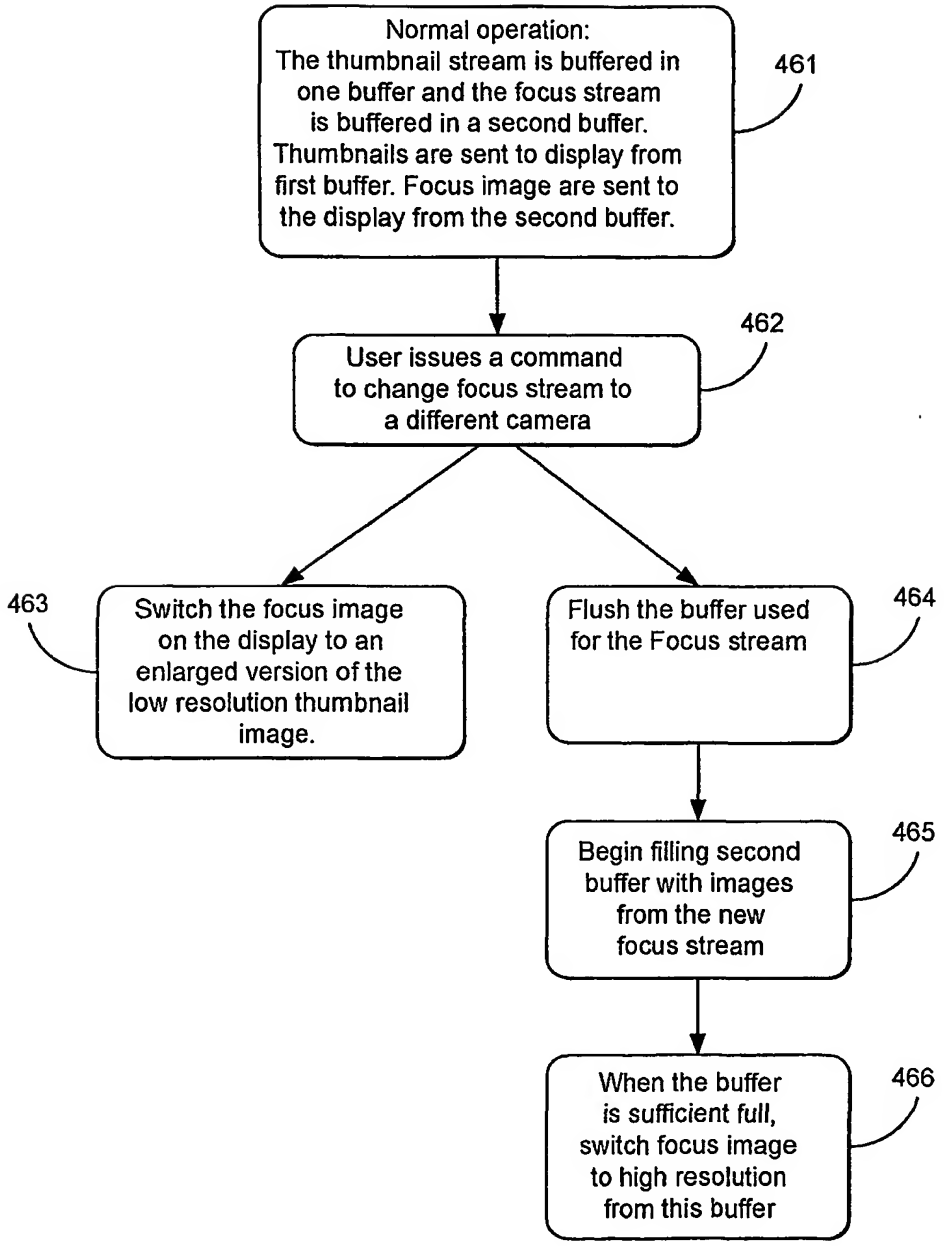


FIG. 4F

(when focus stream changed thumbnail images used initially)

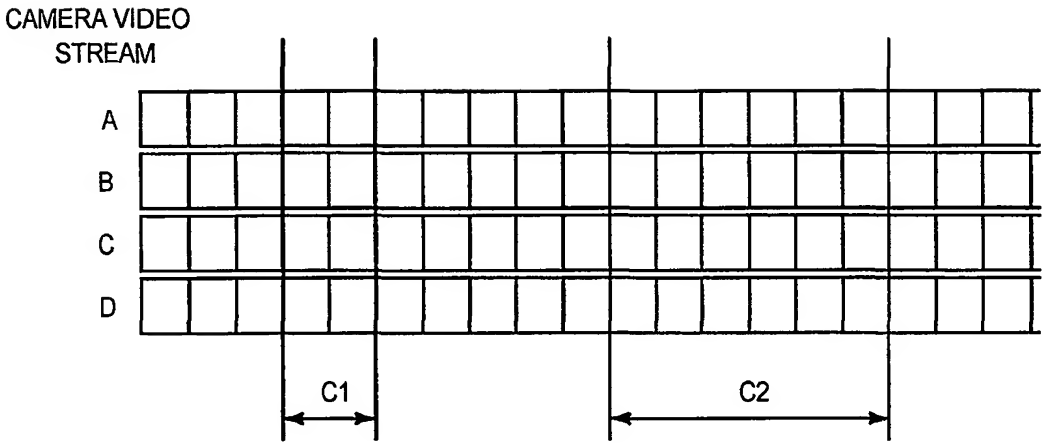


FIG. 5

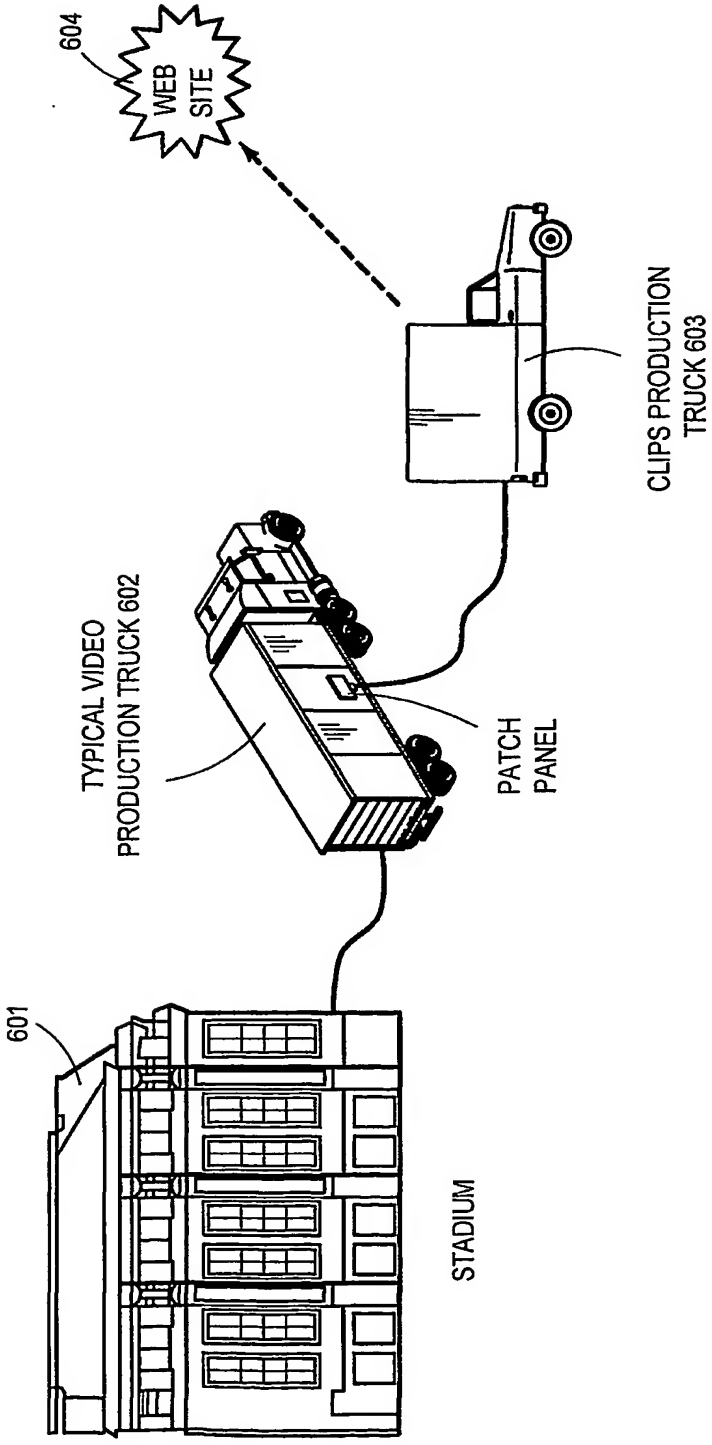

604
WEB SITE
CLIPS PRODUCTION TRUCK 603
TYPICAL VIDEO PRODUCTION TRUCK 602
PATCH PANEL
601
STADIUM

FIG. 6

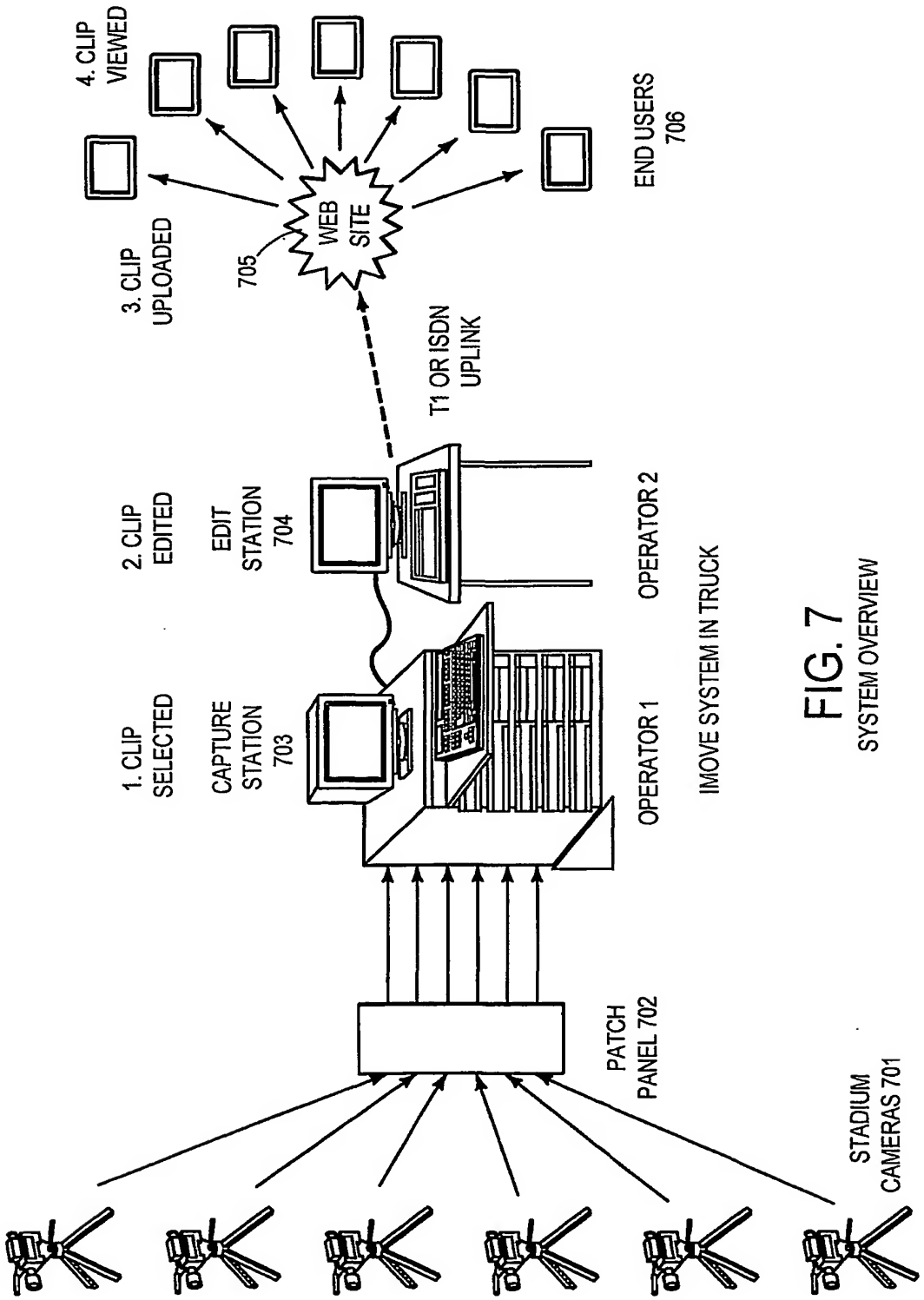


FIG. 7
SYSTEM OVERVIEW

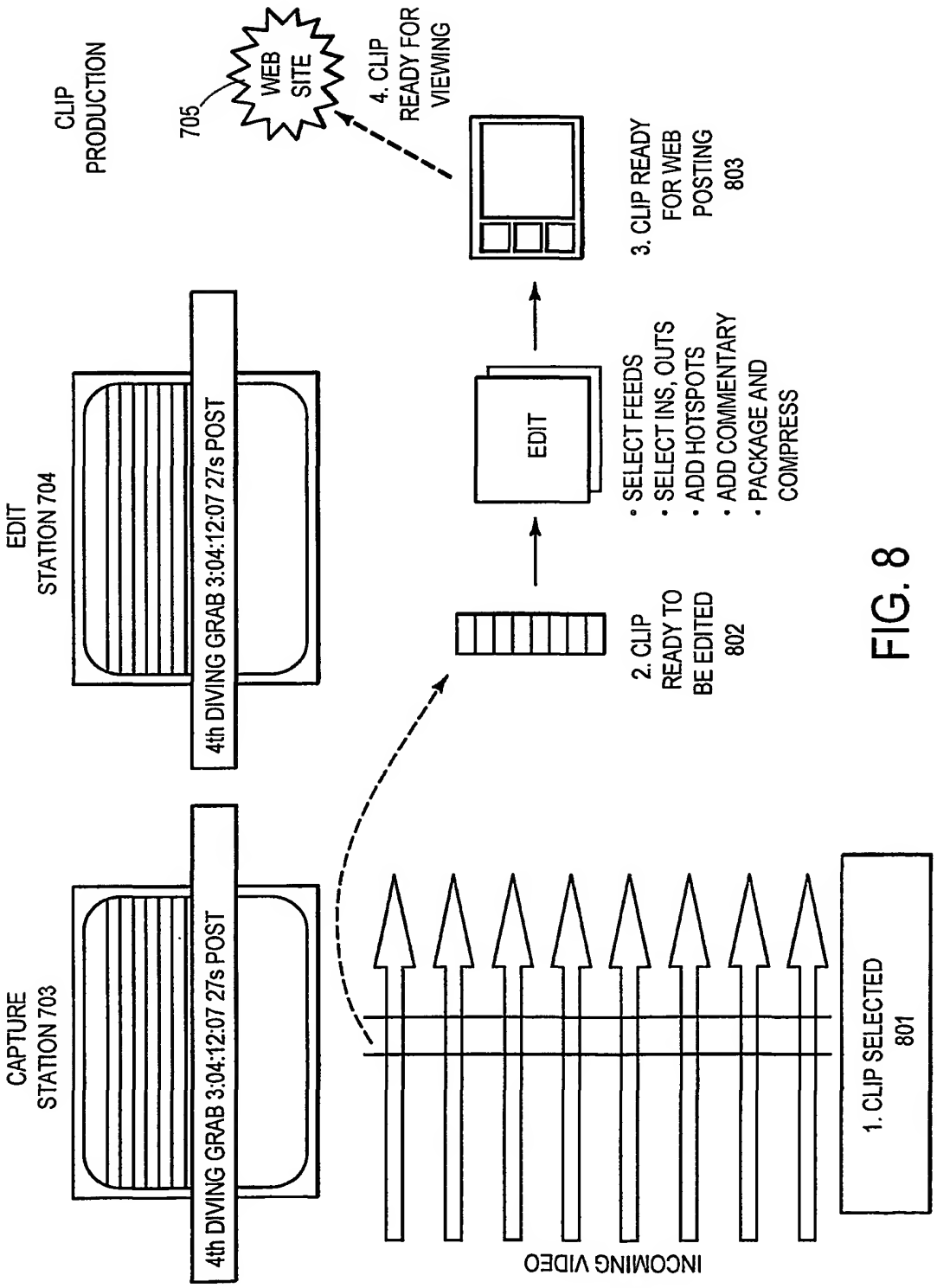

CLIP PRODUCTION
705
WEB SITE
4. CLIP READY FOR VIEWING
3. CLIP READY FOR WEB POSTING 803
EDIT STATION 704
4th DIVING GRAB 3:04:12:07 27s POST
EDIT
• SELECT FEEDS
• SELECT INS, OUTS
• ADD HOTSPOTS
• ADD COMMENTARY
• PACKAGE AND COMPRESS
2. CLIP READY TO BE EDITED 802
CAPTURE STATION 703
4th DIVING GRAB 3:04:12:07 27s POST
1. CLIP SELECTED 801
INCOMING VIDEO

FIG. 8

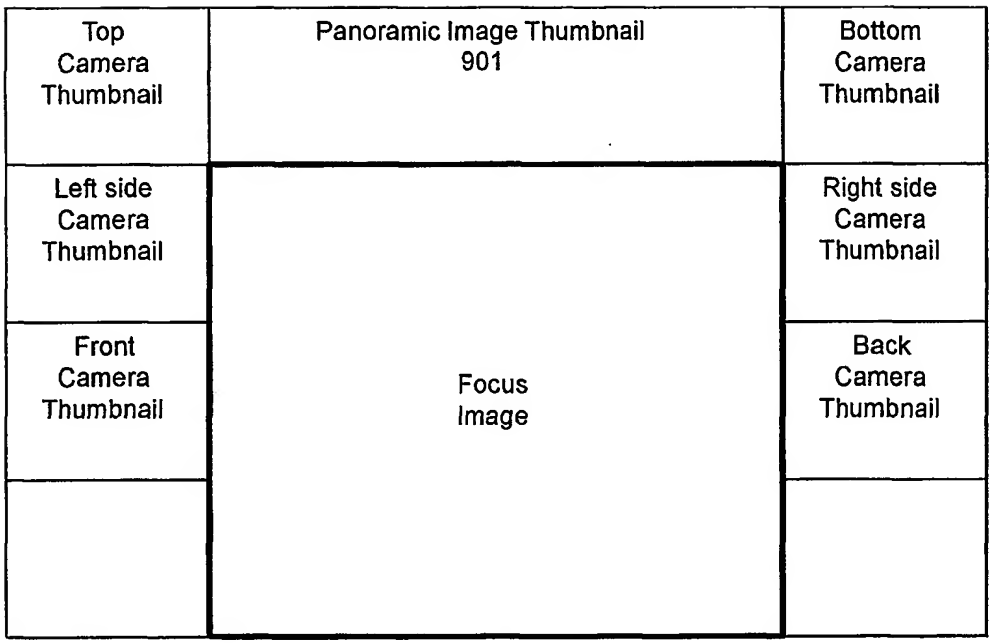

Top Camera Thumbnail
Panoramic Image Thumbnail 901
Bottom Camera Thumbnail
Left side Camera Thumbnail
Right side Camera Thumbnail
Front Camera Thumbnail
Focus Image
Back Camera Thumbnail

FIG. 9

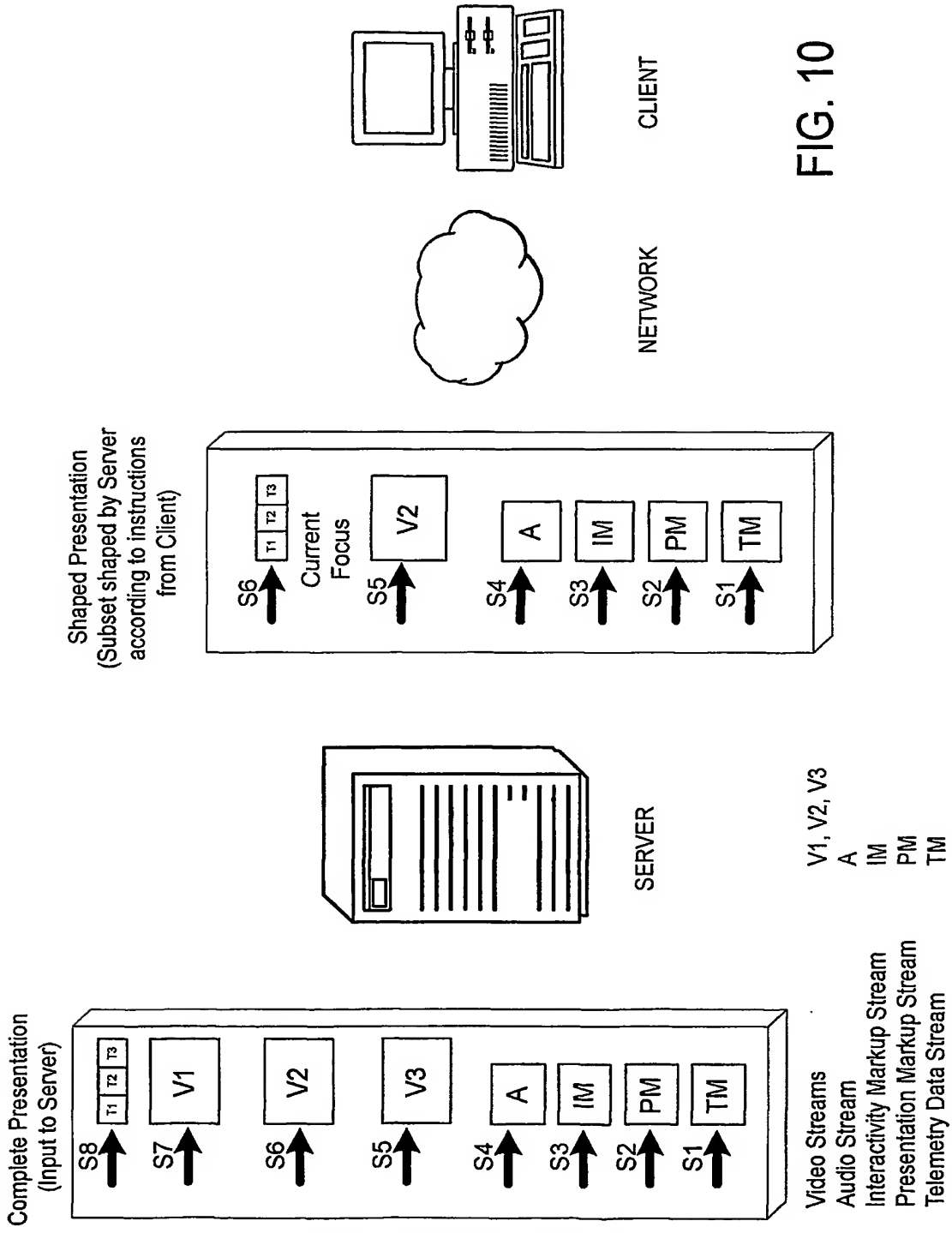

Complete Presentation (Input to Server)
S8
T1
T2
T3
S7
V1
S6
V2
S5
V3
S4
A
S3
IM
S2
PM
S1
TM
Video Streams
Audio Stream
Interactivity Markup Stream
Presentation Markup Stream
Telemetry Data Stream
SERVER
V1, V2, V3
A
IM
PM
TM
NETWORK
CLIENT
Shaped Presentation (Subset shaped by Server according to instructions from Client)
Current Focus

FIG. 10

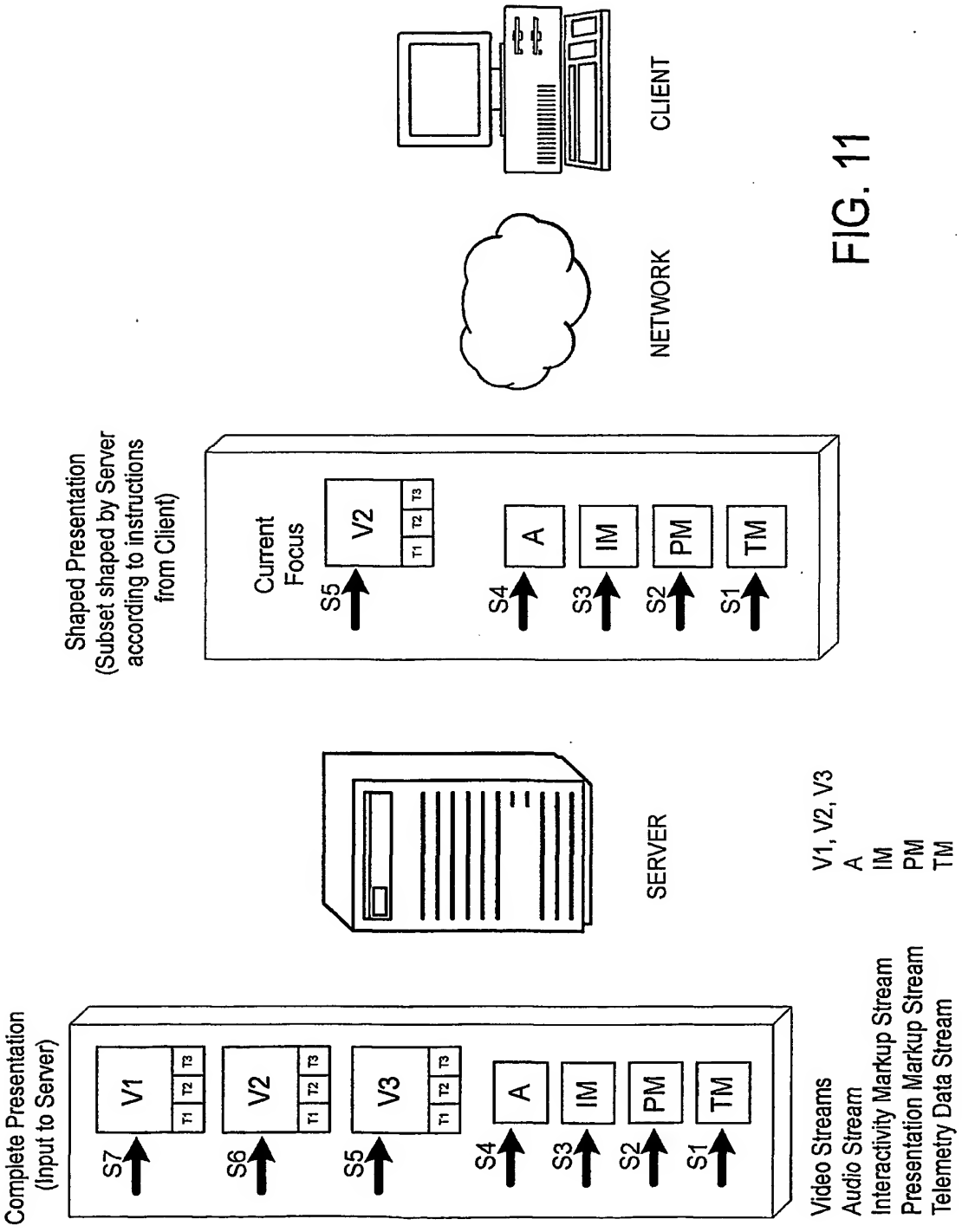

Complete Presentation
(Input to Server)
V1
V2
V3
A
IM
PM
TM
T1
T2
T3
S7
S6
S5
S4
S3
S2
S1
SERVER
Shaped Presentation
(Subset shaped by Server
according to instructions
from Client)
Current
Focus
NETWORK
CLIENT
Video Streams
Audio Stream
Interactivity Markup Stream
Presentation Markup Stream
Telemetry Data Stream
V1, V2, V3
A
IM
PM
TM

FIG. 11

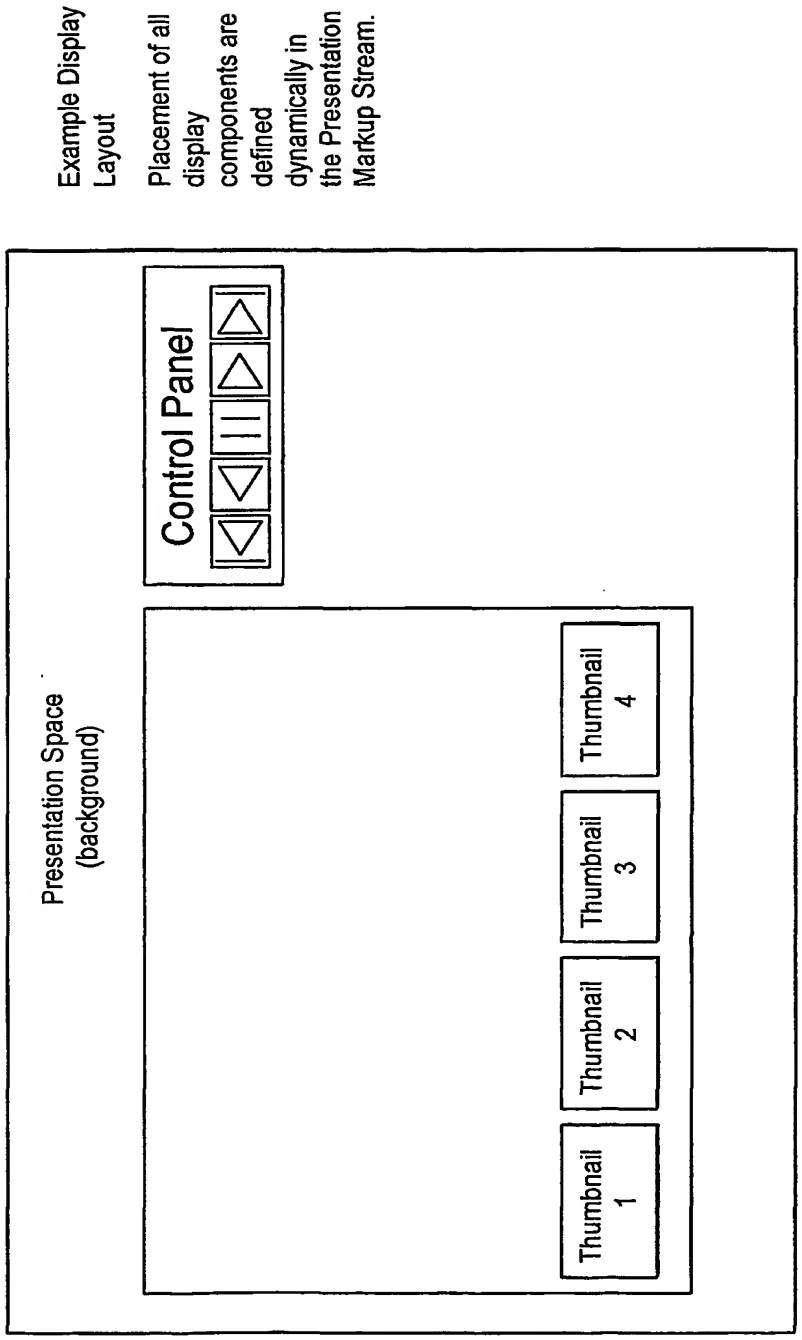

Example Display Layout
Placement of all display components are defined dynamically in the Presentation Markup Stream.
Presentation Space (background)
Control Panel
Thumbnail 1
Thumbnail 2
Thumbnail 3
Thumbnail 4

FIG. 11A

17/17

Interaction between Client and Server over time

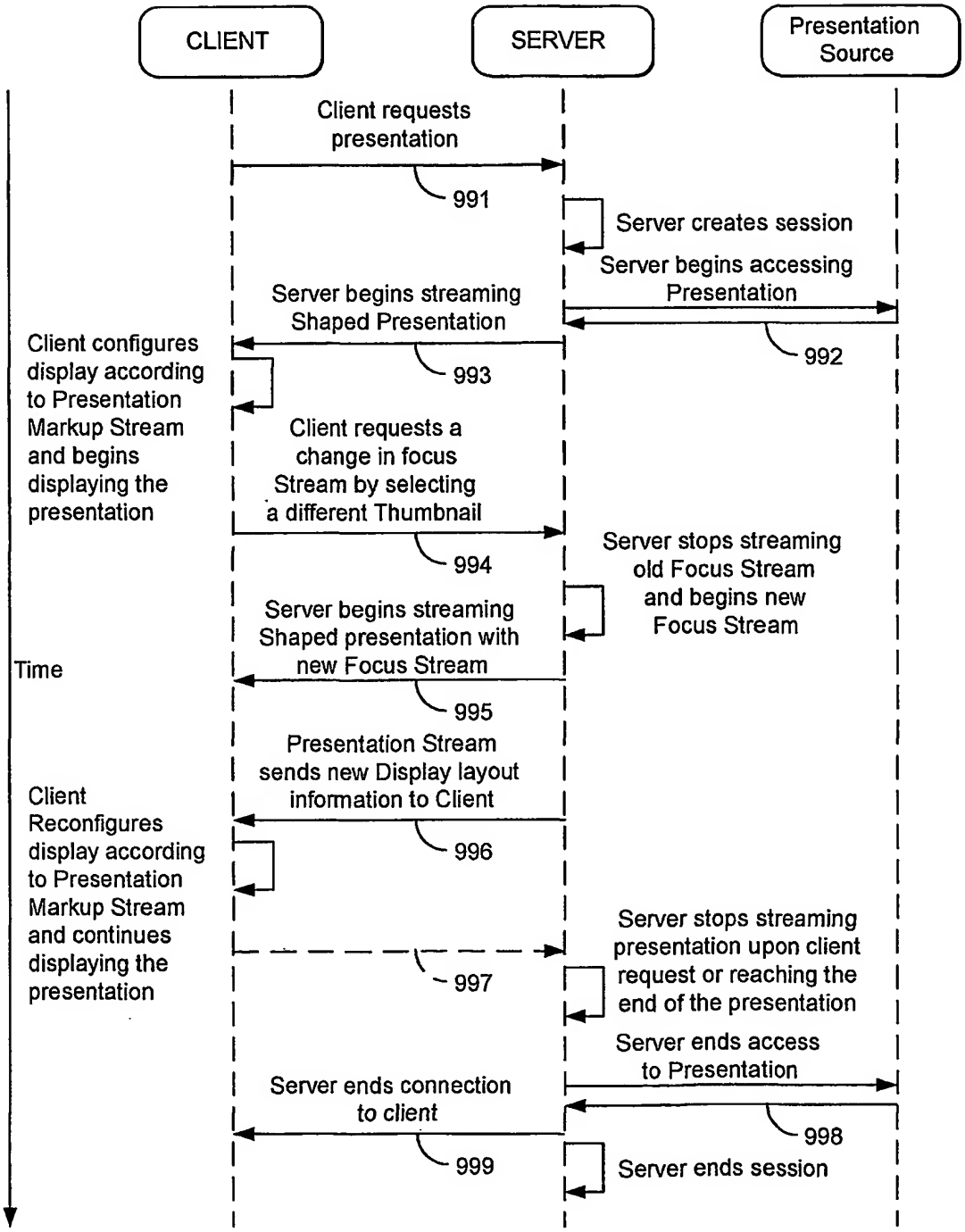


FIG. 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/16289

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :H04N: 07/18
US CL :348/36, 143; 725/105

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 348/36-39, 143-155; 725/105

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS: internet and surveillance

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6,006,265 A (RANGAN et al.) 21 December 1999, column 18, lines 10-50; column 21, lines 40-68; column 23, lines 43-68; column 24, lines 15-68	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier document published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	"&" document member of the same patent family

Date of the actual completion of the international search 08 AUGUST 2001	Date of mailing of the international search report 28 AUG 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ANAND S. RAO Telephone No. (703) 305-4813

Form PCT/ISA/210 (second sheet) (July 1998) ★